The Determinants of VAT Introduction

Cizek, P.; Lei, J.; Ligthart, J.E.

Publication date:
2012

Citation for published version (APA):
Cizek, P., Lei, J., & Ligthart, J. E. (2012). *The Determinants of VAT Introduction: A Spatial Duration Analysis*. (CentER Discussion Paper; Vol. 2012-071). Econometrics.

CentER

Discussion Paper

No. 2012-071

**THE DETERMINANTS OF VAT INTRODUCTION:
A SPATIAL DURATION ANALYSIS**

By

Pavel Čížek, Jinghua Lei, Jenny E. Ligthart

August 29, 2012

ISSN 0924-7815

Tilburg ◆ University

The Determinants of VAT Introduction: a Spatial Duration Analysis

Pavel Čížek* Jinghua Lei† Jenny E. Ligthart‡

August 18, 2012

Abstract

The spatial survival models typically impose frailties, which characterize unobserved heterogeneity, to be spatially correlated. This specification relies highly on a pre-determinate covariance structure of the errors. However, the spatial effect may not only exist in the unobserved errors, but it can also be present in the baseline hazards and the dependent variables. A new spatial survival model with these three possible spatial correlation structures is explored and used to investigate the determinants of value-added tax implementation in 92 countries over the period 1970–2008 using the proposed model. The estimation results suggest the presence of a significant copycat effect among neighboring countries for both contiguity and distance weight matrices.

JEL codes: C11, C23, C41, H20, H70

Keywords: Spatial duration; MCMC; Metropolis-Hastings algorithm; Value-added tax

*Corresponding Author: CentER, Department of Econometrics and OR, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Phone: +31-13-466-8723, Fax: +31-13-466-3280, E-mail: p.cizek@tilburguniversity.edu.

†CentER and Department of Econometrics and Operation Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Phone: +31-13-466-3399, E-mail: j.lei@uvt.nl.

‡CentER, Department of Economics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Phone: +31-13-466-8755, Fax: +31-13-466-4032, Email: j.ligthart@tilburguniversity.edu.

1 Introduction

The value-added tax (VAT), first introduced 50 years ago, remained confined to a few countries until the late 1960s. However, after another 30 years, roughly 150 countries have implemented a VAT, which on average raises about 25 percent of their tax revenue (Ebrill et al., 2001). The VAT is a tax on value added, which can be defined as the value that a producer adds to his raw materials or purchases before selling the improved product or service. Its invoice-credit mechanism—which seeks to tax the value added at each stage of the production-distribution chain—causes it to fundamentally differ from a retail sales tax or a turnover tax. Our aim is to study the factors influencing the introduction of a VAT in a country, and in particular, the dynamic effects of the neighboring countries' (VAT-)decisions on the VAT enactment. For this purpose, we propose a new spatial duration model, discuss its estimation, and apply it to data on VAT adoption covering last 40 years.

Which factors determine the VAT enactment? This question is of pivotal importance to policy makers, but it has received remarkably little attention in the academic literature, especially on the empirical side.[1] Ebrill et al. (2001) provide some informal guidance on selecting potential determinants of VAT adoption.[2] They informally argue that countries are more likely to adopt the VAT if they have a higher GDP per capita, are less open, have a higher literacy rate, and feature a larger population. Recently, Keen and Lockwood (2010) are the first ones to formally explore the causes and consequences of VAT adoption by using a dynamic probit model for a sample of 143 countries during the 1975–2000 period. Their analysis makes a first step in capturing possible neighborhood effects of VAT adoption: countries are more inclined to implement a VAT when other countries in the same region have done so, the so-called copycat effect. However, Keen and Lockwood (2010) neither employ a formal spatial econometric framework nor make use of survival analysis to measure the copycat effect.[3]

[1] Various studies focus on the economic effects of VAT enactment. Nellor (1987) considers empirically the revenue effect of VAT adoption by analyzing a sample of 11 European countries in the 1960s and 1970s and provides evidence that VAT introduction raises the tax revenue-to-GDP ratio. Desai and Hines (2005) examine the effect of VAT implementation on international trade and find that reliance on VAT is associated with less exports and imports. Furthermore, this negative effect on exports is stronger among low-income countries than it is among high-income countries.

[2] We focus on the date of VAT implementation. However, we will refer to adoption, enactment, and implementation interchangeably.

[3] Brockmeyer (2010) uses the Cox proportional hazard model to estimate the impact of lending by the International Monetary Fund (IMF) on VAT adoption in a panel of 125 countries during the period 1975–2000, but she does not focus on the spatial dimension.

To examine this copycat effect, it is not sufficient to incorporate a variable that indicates the proportion of countries with a VAT in the same region. The reason is that these neighborhood effects do not only exist between the observations, they might also occur in the unobserved factors and thus in error term. With regard to the spatial dependence among the observations and in the unobserved component, the traditional (non-spatial) estimation procedures may not be consistent to draw appropriate inferences as their assumptions have been violated. Hence, appropriate inference is not feasible. On the other hand, standard spatial survival model always assume that the spatial correlation structure only exists in the unobserved errors, which is not realistic and does not facilitate examining the spatial correlation explicitly (e.g., see Li and Ryan, 2002; Bastos and Gammerman, 2006).

The purpose of this paper is to develop a new spatial survival model that captures the spatial effects explicitly via an observed spatial lag (i.e., in terms of including a spatially lagged dependent variable) and to apply it to investigation of the neighborhood effects of VAT implementation. To this end, we propose a new spatial survival model, discuss its estimation, and apply it to a unique dataset on VAT adoption spanning the period 1970–2008 (the sample consists of 92 countries, of which 71 countries have adopted a VAT). To capture spatial dependence, our model contains three location-dependent components: (1) spatial dependence exists in the baseline hazard by allowing a region-specific baseline hazard function; (2) spatial dependence is present across the observations via a spatial lag; and (3) spatial dependence occurs in the error terms, which have a distance-based variance-covariance structure. Whereas the first two spatial effects are new in the context of duration models, the last one has been used in various studies as discussed in the following paragraph.

There is a large literature on survival analysis, which studies the time until a specific event takes place. The pioneering work on the proportional hazard model is Cox (1972), which is based on the assumption of proportionality of hazard rates. As this assumption might be too restrictive in practice, models accounting for time-varying covariates (e.g., Gamerman, 1991) and for unobserved heterogeneity were developed. The unobserved heterogeneity is typically captured by means of random effects called frailties; this term and frailty models have been been first introduced by Vaupel et al. (1979). To handle the dependence of unobserved effects in georeferenced data, spatial frailty models were proposed, where the frailties corresponding to certain strata or clusters are spatially arranged (e.g., as clinical sites or geographical regions). Spatial frailty models can be either geostatistical, where the exact geographic locations of the strata are used, or lattice, where only the positions of the strata relative to each other are used (cf. Banerjee et al., 2003). In the geostatistical approach, the frailties are typically modeled as zero-mean Gaussian random variables with a non-diagonal location-based covariance matrix, where the distance correlation matrix has various forms such as exponential, powered exponential, spherical, Matern, and so on (cf. Li and Ryan, 2002). In the lattice approach, the conditionally autoregressive model—which is introduced by Besag (1974) and studied further by Carlin and Banerjee (2002), for instance—is widely used. This type of model studies the discretely indexed regions where the spatial information is based on the adjacency of regions rather than the distance metric.

The proposed model incorporates a spatially weighted dependent variable, which depends on time and is highly correlated with the duration, as well as spatial frailties, which are correlated across cross-sectional units. For the estimation of this model, the Bayesian analysis in the form typically applied to geostatistical duration models and duration models with time-varying covariates is used. The parameters of interest are thus estimated using the Markov chain Monte Carlo (MCMC) technique employed for dynamic survival models by Hemming and Shaw (2002), for instance. In particular, we follow the the MCMC estimation approach of Bastos and Gamerman (2006), who designed it for dynamic survival models with spatial effects. The main difference lies in that Bastos and Gamerman's model allows for time-varying coefficients in the hazard function, whereas our model assumes constant coefficients, but adds the spatial lag of the dependent variable and location-dependent baseline hazard.

Finally, this proposed model and estimation technique is applied to our VAT data. The results provide strong evidence of a copycat effect irrespective of the imposed spatial structure (i.e., based on a contiguity or distance weight matrices). The copycat-effect estimates are found to be quite robust to model specifications with and without frailties.

The remainder of the paper is organized as follows. In Section 2, a detailed description of the proposed spatial duration model and its estimation is provided. In Section 3, we present the dataset, describe the MCMC simulation setup, and discuss empirical results. Finally, the conclusion follows in Section 4.

2 The Model

In this section, the proposed spatial duration model is introduced, the estimation procedure is described, including the choices of prior distributions of parameters, and the posterior inference and computation are discussed.

2.1 The Spatial Duration Model

Consider a sample of $i = 1, ..., N$ units (e.g., countries) observed for $t = 1, ..., T$ years. Let $X_{i,t-a}$ be a k-dimensional vector of duration-dependent covariates measured at time $t - a$ for unit i, where k is the number of variables in $X_{i,t-a}$ and $a \geq 0$ is a suitable lag. The group of duration-independent covariates with l variables is denoted by Z_i. Further, let $y_{i,t-b}$ be the event (survival) dummy, which equals unity if the analyzed event (e.g., the VAT introduction) occured in unit i at time $t - b$ or earlier and is zero otherwise; $b \geq 0$ denotes again a suitable lag. It is straightforward to see that $y_{i,t-b}$ is duration dependent as it could also be defined as an indicator function $I(t_i \leq t - b)$. The choices of lags a and b are application specific. In our case, because there is a time lag between the time of negotiation, adoption by parliament, and the actual implementation of a VAT, the integer parameters a and b will be used to capture the time lags specific to the VAT enactment (see Section 3.1), but they can also equal 0 in other applications.

The proportional hazard function of the proposed spatial duration model is given by

$$\lambda_i\left(t\right) = \lambda_0\left(t, s_i\right) \exp\left[X_{i,t-a}^{\top}\beta + Z_i^{\top}\gamma + \rho W_i y_{\cdot,t-b} + U_i(s_i)\right], \tag{1}$$

where $\lambda_0\left(t, s_i\right)$ is the baseline hazard rate dependent on the duration t and the spatial location s_i of unit i. As in the standard proportional hazard models (Cox, 1972), the hazard rate depends on the parameters β and γ of the covariates. The spatial extension of the model consists of three elements. First, the baseline hazard $\lambda_0\left(t, s_i\right)$ is allowed to be a function of location s_i. Next, the spatial interaction term $\rho W_i y_{\cdot,t-b}$ consist of the parameter ρ and the spatially lagged values of the event dummy $W_i y_{\cdot,t-b}$, where W_i refers to the ith row of the spatial weight matrix W_N (defined and discussed below) and $y_{\cdot,t-b} = (y_{1,t-b}, \ldots, y_{N,t-b})^{\top}$. Finally, the spatial frailty $U_i(s_i)$ is a second-order stationary zero-mean process, that is, $\mathrm{E}\left[U_i(s_i)\right] = 0$, $\mathrm{Var}\left[U_i(s_i)\right] = \sigma^2$, and $\mathrm{Cov}\left[U_i(s_i), U_j(s_j)\right] = \sigma^2\varphi(s_i, s_j; \phi)$ for all $i \neq j$, where $\varphi(s_i, s_j; \phi)$ is a valid two-dimensional correlation function (see Section 2.3), ϕ is a parameter controlling this function, and σ^2 denotes the variance of frailties. This setting represents a traditional approach in geostatistical modeling to capture spatial association among observations at a fixed set of spatial locations s_i.

Our model contains many common models as special cases: (i) the Cox (1972) model is obtained if $\lambda_0(t, s_i) = \lambda_0(t)$ and $\gamma = \rho = U_i = 0$; (ii) the frailty model (Clayton, 1978) is derived if $\lambda_0(t, s_i) = \lambda_0(t)$ and $\gamma = \rho = \phi = 0$; (iii) the frailty model with group-varying baseline hazards (Carlin and Hodges, 1999) is obtained if $\gamma = \rho = \phi = 0$; (iv) spatial frailty models (Carlin and Banerjee, 2002) are obtained if $\lambda_0(t, s_i) = \lambda_0(t)$ and $\gamma = \rho = 0$. (Note that $\phi = 0$ is assume to correspond to $\varphi(s_i, s_j; \phi) = 0$ irrespective of s_i and s_j.)

The spatial weight matrix W_N has to be specified to estimate the model as it describes which neighbors of unit i influence its hazard rate. Typically, W_N is assumed to be row-normalized so that W_i in (1) contains weights summing up to 1. In the VAT application, we use two different spatial weight matrices. The first one is the contiguity matrix, which only indicates whether two countries share a common border. The elements of this weight matrix are $w_{ij}^C = b_{ij}/\sum_{i=1}^N b_{ij}$ for $i \neq j$ and $w_{ij}^C = 0$ for $i = j$, where b_{ij} is the border dummy that equals to one if countries i and j share a common border and zero otherwise. The second kind of the spatial weight matrix employs the inverse of the squared distance between the largest cities of two countries to reflect the gravity type of approach.[4] In contrast to the previous measure, the distance scheme enables copycat effects to exist among all countries. The elements of this distance matrix can be characterized by:

$$w_{ij}^D = \begin{cases} d_{ij}^{-2}/\sum_{j=1}^N d_{ij}^{-2} > 0 & \text{for } i \neq j \\ 0 & \text{for } i = j \end{cases}, \tag{2}$$

where d_{ij} reflects the geographical distance between the largest cities of countries i and j computed as the great circle distance given latitude and longitude.

Finally, the specification of the baseline hazard function $\lambda_0(t, s_i)$ is discussed since it is specified only as a general function of time and location. While this general specification requires just some additional regularity assumptions to be identifiable and estimable, the focus of this paper is on a piecewise-constant baseline hazard function defined on a one- or

[4]Other alternatives to measure the distance between two countries can be based on the amount of trade between two countries, for instance.

two-dimensional grid because of the scarcity of data and ease of estimation. Additionally in large data sets, this piecewise-constant approach can be used as a nonparametric estimator of the baseline hazard function if the grid size is let to converge to 0 with an increasing sample size. Assuming first that the baseline hazard function depends only on time, $\lambda_0(t, s_i) = \lambda_0(t)$, the time span can be partitioned into the following disjoint intervals following the specification of Gamerman (1991):

$$I_j = \begin{cases} [0, \alpha_1] & j = 1 \\ (\alpha_{j-1}, \alpha_j] & j = 2, ..., J-1 \\ (\alpha_{J-1}, +\infty) & j = J \end{cases}$$

with $\alpha_0 = 0 < \alpha_1 < ... < \alpha_J < +\infty$. The baseline hazard rate is then given by

$$\lambda_0(t) = \lambda_j \quad \text{if} \quad t \in I_j, \tag{3}$$

where $\lambda_j > 0$ for $j = 1, ..., J$. An extension of model (3) allows the baseline hazard rates to differ across regions, but to be constant within each region (a region is assumed to be a set of neighboring locations; for example, a continent can represent a region when units are countries). More specifically, the baseline hazard rate in region $r = 1, ..., R$ is given by $\lambda_{j,r}$ if duration t belongs to I_j, where $\lambda_{j,r} > 0$ for all r and j. If R_{ir} denotes the regional dummies, where $R_{ir} = 1$ if unit (country) i is in region r and zero otherwise, the baseline hazard can be defined as

$$\lambda_0(t, s_i) = \sum_{j=1}^{J} \sum_{r=1}^{R} \lambda_{j,r} I(t \in I_j) I(R_{ir} = 1) = \sum_{j=1}^{J} \sum_{r=1}^{R} \lambda_{j,r} I(t \in I_j) R_{ir}. \tag{4}$$

2.2 Maximum Likelihood Estimation

Consider now the survival sample of size N, $D = \{t_i, \delta_i, s_i, X_{i1}, \ldots, X_{i,t_i}, Z_i, y_{i,1}, \ldots, y_{i,t_i}\}_{i=1}^{N}$, where t_i denotes the duration (e.g., time to the VAT introduction) and δ_i is the right-censoring indicator: $\delta_i = I(\text{'no censoring'})$. This sample should be used to estimate the parameters of model (1): the parameter vector $\theta = \left[\ln \lambda_0^\top, \beta^\top, \gamma^\top, \rho\right]^\top$, where the baseline hazard function is uniquely defined by JR parameter values $\lambda_0 = \{\lambda_{j,r} : j = 1, \ldots, J; r = 1, \ldots, R\}$, see (4). As the estimation of model (1) relies on the maximum likelihood estimation, the likelihood function has to be constructed first.

Let T denote the (random) duration without censoring and $F\left(t|\mathcal{I}_{i,t},\theta\right)$ and $f\left(t|\mathcal{I}_{i,t},\theta\right)$ describe the conditional distribution and density functions of T, respectively, where the information set $\mathcal{I}_{i,t}=[s_i,X_{i,1},\ldots,X_{i,t-a},Z_i,y_{.,t-b},U_i(s_i)]$ represents all information observed prior to time t. For the right-censored observations, we only know that the durations exceed t, so the complement of the cumulative distribution function is

$$S\left(t|\mathcal{I}_{i,t},\theta\right)=\mathrm{P}\left(T>t|\mathcal{I}_{i,t},\theta\right)=\int_t^\infty f\left(v|\mathcal{I}_{i,t},\theta\right)dv=1-F\left(t|\mathcal{I}_{i,t},\theta\right), \tag{5}$$

where $S(.)$ denotes the survival function. Consequently, the likelihood contribution of the ith observation can be written as

$$f\left(t_i|\mathcal{I}_{i,t},\theta\right)^{\delta_i}S\left(t_i|\mathcal{I}_{i,t},\theta\right)^{1-\delta_i},$$

where δ_i is the right-censoring indicator. Hence, the conditional likelihood function equals

$$L\left(\theta\right)=\prod_{i=1}^N f\left(t_i|\mathcal{I}_{i,t},\theta\right)^{\delta_i}S\left(t_i|\mathcal{I}_{i,t},\theta\right)^{1-\delta_i}. \tag{6}$$

We first derive the likelihood for the case that the baseline hazard depends on the duration, but not on the location: $\lambda_0(t,s_i)=\lambda_0(t)$. For simplicity, the duration-dependent (time-varying) components will be now denoted by $V_{i,t}(\theta)=X_{i,t-a}^\top\beta+\rho W_i y_{.,t-b}$ and the duration-independent (time-constant) components by $C_i(\theta)=Z_i^\top\gamma+U_i(s_i)$. Since generally the hazard function $\Lambda=-\ln S$, the hazard rate $\lambda=\Lambda'$, and the density function $f=-S'$, it follows that $\ln f=\ln[\lambda S]=\ln\lambda+\ln S$ and the conditional likelihood function can be written as

$$\begin{aligned} L(\theta) &= \exp\left\{\sum_{i=1}^N\left[\delta_i\ln\lambda(t_i|\mathcal{I}_{i,t},\theta)+\ln S(t_i|\mathcal{I}_{i,t},\theta)\right]\right\} \\ &= \exp\left\{\sum_{i=1}^N\left[\delta_i\left(\ln\lambda_0(t_i)+C_i(\theta)+V_{i,t_i}(\theta)\right)-\exp[C_i(\theta)]\int_0^{t_i}\lambda_0(v_i)\exp[V_{i,v_i}(\theta)]dv_i\right]\right\}. \end{aligned} \tag{7}$$

As in Gupta (1991), we assume that all covariates stay constant for finite subperiods of time, that is, $V_{i,t}(\theta)$ stays constant in the duration interval t to $t+1$ and jumps to $V_{i,t+1}(\theta)$

at period $t+1$. Therefore, the integral in equation (7) can be expressed as

$$\int_{0}^{t_i} \lambda_0(v_i) \exp\left[V_{i,v_i}(\theta)\right] dv_i = \sum_{v_i=0}^{t_i} \lambda_0(v_i) \exp\left[V_{i,v_i}(\theta)\right]. \tag{8}$$

For the proposed piecewise-constant baseline hazard (3), the conditional likelihood function can be finally rewritten as

$$\begin{aligned} L(\theta) \quad = \quad & \exp\left\{ \sum_{i=1}^{N} \left[\delta_i \left(\sum_{j=1}^{J} [I(t_i \in I_j) \ln \lambda_j] + C_i(\theta) + V_{i,t_i}(\theta) \right) \right.\right. \\ & \left.\left. - \exp\left[C_i(\theta)\right] \sum_{j=1}^{J} [I(t_i \in I_j) D_j^T(t_i, \theta)] \right] \right\}, \end{aligned} \tag{9}$$

where

$$D_j^T(t_i,\theta) = \begin{cases} \sum\limits_{v_i=\alpha_0}^{t_i} [\lambda_1 \exp(V_{i,v_i}(\theta))] & t_i \in I_1, \\ \sum\limits_{k=1}^{j} \sum\limits_{v_i=\alpha_{k-1}}^{\alpha_k} [\lambda_k \exp(V_{i,v_i}(\theta))] + \sum\limits_{v_i=\alpha_{j-1}}^{t_i} [\lambda_j \exp(V_{i,v_i}(\theta))] & t_i \in I_j, j = 2, ..., J. \end{cases}$$

In the more general case (4), when the baseline hazard rates are allowed to differ across regions, an analog to equation (9) results in the conditional likelihood function

$$\begin{aligned} L(\theta) \quad = \quad & \exp\left\{ \sum_{i=1}^{N} \left[\delta_i \left(\sum_{r=1}^{R} \sum_{j=1}^{J} [I(t_i \in I_j) R_{ir} \ln \lambda_{j,r})] + C_i(\theta) + V_{i,t_i}(\theta) \right) \right.\right. \\ & \left.\left. - \exp\left[C_i(\theta)\right] \sum_{r=1}^{R} \sum_{j=1}^{J} [I(t_i \in I_j) R_{ir} D_{j,r}^{TR}(t_i, \theta)] \right] \right\}, \end{aligned} \tag{10}$$

where

$$D_{j,r}^{TR}(t_i,\theta) = \begin{cases} \sum\limits_{v_i=\alpha_0}^{t_i} [\lambda_{1,r} \exp(V_{i,v_i}(\theta))] & t_i \in I_1, \\ \sum\limits_{k=1}^{j-1} \sum\limits_{v_i=\alpha_{k-1}}^{\alpha_k} [\lambda_{k,r} \exp(V_{i,v_i}(\theta))] + \sum\limits_{v_i=\alpha_{j-1}}^{t_i} [\lambda_{j,r} \exp(V_{i,v_i}(\theta))] & t_i \in I_j, j = 2, ..., J. \end{cases}$$

As the likelihood functions described in this section do not yield a closed-form solution, owing to the unobserved frailties, the standard likelihood maximization cannot be applied.

Therefore, in the context of spatial survival analysis, the Bayesian estimation in combination with the MCMC method is used to obtain an approximation of the posterior distribution of the parameters through posterior samples (cf. Hemming and Shaw, 2002; Bastos and Gamerman, 2006). For such analysis, the prior distributions of all parameters have to be specified in Section 2.3. The posterior distributions of the parameters and their approximation by means of the MCMC method are described in Section 2.4.

2.3 Prior Distribution

The spatial duration model (1) contains parameters of two groups: regression coefficients θ and spatial frailties U. Let us now specify the prior distribution of these two parameter groups, which are assumed to be conditionally independent. The parameters of these two prior distributions, the so-called hyper-parameters, are denoted by Ψ_θ and Ψ_U, respectively. The complete set of hyper-parameters is referred to by $\Psi = (\Psi_\theta, \Psi_U)$ and is assumed to be independent of the prior distributions of regression coefficients and spatial frailties. Therefore, the full prior satisfies

$$p(\theta, U, \Psi) = p(\theta|\Psi_\theta)p(U|\Psi_U)p(\Psi).$$

The prior density that we assign to θ is the normal prior density. Hence, the means b_θ and variances T_θ for these parameters will have to be specified. As $\Psi_\theta = (b_\theta, T_\theta)$, the prior specification can be expressed as follows:

$$\theta \sim N(b_\theta, T_\theta) \Longleftrightarrow p(\theta|\Psi_\theta) \propto \exp\left\{-(1/2)(\theta - b_\theta)^\top T_\theta^{-1}(\theta - b_\theta)\right\}. \tag{11}$$

Concerning the spatial frailties, there is a number of possibilities to model the geographical correlation when introducing the spatial dependence between observations (see Li and Ryan, 2002) that are shown to be valid in the sense that the resulting variance-covariance matrices are positive definite on some open parameter sets (Ripley, 1981). Bastos and Gamerman (2006) argue that the Gaussian process approach is flexible and accommodates many different forms of spatial dependence, which are easily and directly implemented through the correlation function. Furthermore, this approach could be easily applied to data based both on continuously distributed locations and discrete disjoint regions.

In our spatial duration context, we apply the Gaussian correlation function with $k = 1$, which is also called the exponential correlation function as the correlation function introducted in Section 2.1 equals $\varphi(s_i, s_j; \phi) = \exp(-||s_i - s_j||/\phi)$ ($||s_i - s_j||$ denotes here the Euclidean distance between the locations of units i and j). Therefore, the joint density of the spatial frailties is given by

$$P(U|\Psi_U) = p(U|\sigma^2, \phi) \propto (\sigma^2)^{-n/2}|H(\phi)|^{-1/2} \exp[-\frac{1}{2\sigma^2}U^\top H^{-1}(\phi)U],$$

where $H(\phi)_{ij} = \varphi(s_i, s_j; \phi) = \exp(-||s_i - s_j||/\phi)$ for $i, j = 1, \ldots, N$. Following Bastos and Gamerman (2006), the prior of hyper-parameter σ^2 is assumed to be Inverse Gamma (IG) distributed, $IG(\frac{a_\sigma}{2}, \frac{b_\sigma}{2})$, and the prior of the hyper-parameter ϕ is assumed to be Gamma distributed, $G(a_\phi, b_\phi)$.

2.4 Posterior Inference and Computation

The posterior distribution is obtained by combining the likelihood function with the prior distribution according to the Bayesian theorem. For example, the posterior density for model (9) in Section 2.1 is, conditional on observed data, given by

$$\begin{aligned} p(\theta, U, \Psi|D) \quad &\propto \quad \exp\left\{\sum_{i=1}^{N}\left[\delta_i\left(\sum_{j=1}^{J}[I(t_i \in I_j)\ln\lambda_j] + C_i(\theta) + V_{i,t_i}(\theta)\right)\right.\right. \\ &\qquad\qquad \left.\left. -\ \exp\left[C_i(\theta)\right]\sum_{j=1}^{J}[I(t_i \in I_j)D_j^T(t_i, \theta)]\right]\right\}, \\ &\times \quad \exp\{-\frac{1}{2}(\theta - b_\theta)^\top T_\theta^{-1}(\theta - b_\theta)\} \\ &\times \quad (\sigma^2)^{-n/2}|H(\phi)|^{-1/2}\exp\{-\frac{1}{2\sigma^2}U^\top H^{-1}(\phi)U\} \\ &\times \quad (\sigma^2)^{-\frac{a_\sigma}{2}-1}\exp(-\frac{b_\sigma}{2\sigma^2})\phi^{a_\phi - 1}\exp(-b_\phi\phi) \end{aligned}$$

(cf. Bastos and Gammerman, 2006, eq. (23)); the expression for model (10) looks analogously.

This posterior distribution does not possess a closed-form analytic solution. Therefore, MCMC methods will be used to approximate it through posterior sampling. As the conditional distributions can be expressed mathematically, but do not have an explicit form, Gibbs sampling (Geman and Geman, 1984) is difficult. Similarly to Bastos and Gamerman (2006),

Table 1: Sampling Algorithm

Step 1: Specify an initial value $\psi^{(0)} = (\theta^{(0)}, \sigma^{(0)}, U^{(0)}, \phi^{(0)})$ and set $j = 1$.
Step 2: Repeat for $j = 1, 2, ..., M$. (1) $\theta^{(j)} \sim p(\theta \vert U^{(j-1)}, (\sigma^2)^{(j-1)}, \phi^{(j-1)}, D)$ (2) $U^{(j)} \sim p(U \vert \theta^{(j)}, (\sigma^2)^{(j-1)}, \phi^{(j-1)}, D)$ (3) $(\sigma^2)^{(j)} \sim p(\sigma^2 \vert \theta^{(j)}, U^{(j)}, \phi^{(j-1)}, D)$ (4) $\phi^{(j)} \sim p(\phi \vert \theta^{(j)}, U^{(j)}, (\sigma^2)^{(j)}, D)$
Step 3: Return the values $\{\psi^{(1)}, \psi^{(2)}, ..., \psi^{(M)}\}$.

we therefore use the random walk Metropolis-Hastings algorithm, which generates a sequence of draws from the joint posterior distribution of parameters (Metropolis et al., 1953; Hastings, 1970). This procedure is capable of constructing an optimal proposal distribution and can be used to generate samples from an arbitrary density. Additional statistical inference can be carried out when the samples are obtained. The sampling algorithm is described in Table 1.

A key factor in achieving high efficiency of the Metropolis-Hastings algorithm is finding a good proposal distribution for the parameters, where the proposal distribution should resemble the true posterior distribution of parameters. As large sample theory states that the posterior distribution of the parameters approaches a multivariate normal distribution, a normal proposal distribution often works well in applications (Gelman et al., 2004).

The sampling efficiency of the Metropolis-Hastings chain is closely related to the acceptance rate. Roberts et al. (1997) show for the random walk Metropolis-Hastings algorithm that the optimal acceptance probability should be around 0.45 for a single parameter and approach 0.23 in higher dimensions. If the acceptance rate is high and thus most new samples occur right around the current data points, the Markov chain is moving rather slowly, the parameter space cannot be fully explored, and the autocorrelation of the chain is also high. On the other hand, if the acceptance rate is low and thus the proposed samples are often rejected, the chain is not moving much and the variability of the chain will be underestimated. Although it is almost impossible to have the exact desired acceptance rate for a Metropolis chain, Roberts and Rosenthal (2001) empirically demonstrate that acceptance rates between 0.15 and 0.5 achieve at least 80 percent efficiency. Therefore, the acceptance probability could be within a small tolerable range of the optimal values.

3 Empirical Application

In this section, our dataset is described, the setup of the MCMC algorithm is discussed, and the estimation results are provided.

3.1 VAT Data

We apply the methodology introduced in Section 2 to the analysis of VAT adoption. The potential sample is an unbalanced panel of 131 countries that have adopted a VAT over the period 1970–2008. The date of VAT implementation is obtained from data files of the IMF's Fiscal Affairs Department, the Tax News Service of the International Bureau of Fiscal Documentation, and various other sources. The effective sample size in the benchmark case is only 92 countries, reflecting missing data (see below). The list of countries, descriptive statistics, and data definitions with data sources are reported in Tables A.1–A.3, respectively.

Countries of the former Soviet Union and Eastern Europe were dropped from the full sample because they faced several wider structural reforms—including downsizing of the public sector—at the time of VAT introduction. As a result, a negative correlation between public revenue and VAT adoption may be introduced. Furthermore, no reliable data—except artificially constructed data—were available for the pre-1992 period as these countries did not exist yet. The resulting subsample contains 108 countries.

Table A.4 reports missing data in the subsample. (All variables are described in Section 3.2 and Table A.3.) The missing data concentrate themselves in the series of the government revenue-to-GDP ratio, which is denoted by R^q for $q = \{A, B\}$. A new series REV is constructed using R^A and R^B, which are two revenue-to-GDP ratios obtained from an official and internal IMF data source, respectively. The new series has less missing values.[5] In order to obtain a balanced panel without gaps—which is required by our MCMC algorithm—the

[5] Although R^A and R^B are derived from different sources—which employ slightly different definitions of public revenue and have different missing years—they are observed to be proportional to each other. Suppose there is a constant ratio Q such that $\mathrm{E}(R^A) = Q * \mathrm{E}(R^B)$, we define $REV = \frac{R^A+R^B}{2}$ as the new series. In this case, $\mathrm{E}(REV) = \mathrm{E}(\frac{R^A+R^B}{2}) = \frac{Q+1}{2}\,\mathrm{E}(R^B) = \frac{1}{2}(1+\frac{1}{Q})\,\mathrm{E}(R^A)$. Thus, when both R^A and R^B are not missing, REV is the average of these two; if R^A is missing but R^B is not missing, $REV = \frac{Q+1}{2}R^B$; if R^A is not missing but R^B is missing, $REV = \frac{Q+1}{2}R^A$.

Table 2: VAT Adoption by Region Across Various Time Periods

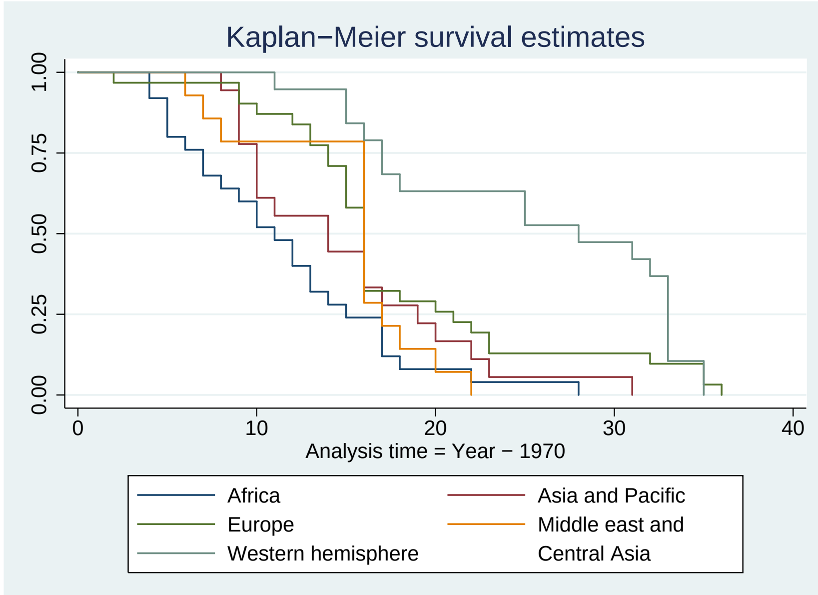
Year	MECA	EU	WH	AP	AF	Total
Total	7	10	12	17	25	71
2001–2008	2	0	0	0	8	10
1996–2000	1	0	1	8	7	17
1991–1995	1	2	4	4	7	18
1986–1990	3	3	0	3	2	11
1981–1985	0	2	2	1	0	5
1976–1980	0	0	2	1	1	4
1970–1975	0	3	3	0	0	6

Notes: Figures show the number of countries adopting a VAT during the particular time frame. MECA, EU, WH, AP, and AF denote Middle East and Central Asia, Europe, Western Hemisphere, Asia and Pacific, and Africa, respectively.

missing data are interpolated using several deterministic rules. First, a missing data point is replaced by the value obtained from its nearest neighborhood. An average of two data points is applied if there are two neighboring values. Otherwise, linear interpolation between the nearest preceeding and following observations in time is used. Second, the values after VAT adoption are never used for interpolating missing values before VAT adoption. Third, the maximum allowed number of consecutive missing values imputed by linear interpolation is set to three. If we stick to the first three rules and then drop the countries with missing data, 52 countries remain, which yields our small sample. However, if we allow linear interpolation when the missing data gap is more than three, 92 countries are left. As there are no qualitative differences between the estimates for the two samples, the latter one containing 92 countries is our main sample.

The mean duration before VAT introduction is 25.3 years. More detailed information is reported in Table 2, which shows VAT adoption by region and time period. Once a VAT is introduced, it stays in place, suggesting that there is a lot of sluggishness in the VAT legislation. Only two countries (i.e., Belize and Malta) repealed the VAT and thus exited the sample. Further, Figure 1 shows the Kaplan-Meier survival estimates of duration times in five different regions. Survival rates are the highest in the Middle Eastern region (see the right-hand side of figure). It can also be seen that the VAT adoption speed is varying in different time intervals. Therefore in our final model, we distinguish the following four intervals in

Figure 1: Kaplan-Meier Survival Estimates



definitions (3) and (4): 1970–1990, 1991–2000, 2001–2008, and 2008–(censored). The survival rates in these intervals seem to be different as indicated in Figure 1. Note that finer grids with more time intervals were considered as well, but the effects of explanatory variables on the hazard rate did not change.

Concerning the employed explanatory variables, there is typically a lag between the time a country starts contemplating adopting a VAT and its actual adoption. Ebrill et al. (2001) describe it takes roughly 18 months from the initial preparations until the passage by parliament of the VAT law. We therefore apply a two-year time lag to the covariates and set the parameters a and b introduced in Section 2.1 equal to 2.

The treatment of the initial observations is an important theoretical and practical problem in dynamic nonlinear panel data models. Much attention has been devoted to dynamic panel data models with unobserved individual fixed effects. Although our model also theoretically faces such an initial observation problem, it is not a real concern since the unobserved heterogeneity captured by frailties is, conditionally on observables, only a function of locations rather than of the lagged dependent dummy variables. Moreover, our panel data cover years from 1970 to 2008. If no countries had adopted a VAT before 1970, then the initial condition problem would be easily solved by simply assuming $y_0 = 0$ for all countries. However, only eight countries in our dataset—i.e., Ivory Coast (1960), Brazil (1967), Denmark (1967), France

(1968), Uruguay (1968), Germany (1968), the Netherlands (1969), and Sweden (1969)—have adopted a VAT before 1970.[6] Because we do not have control variables for the pre-1970 period, we do not use the VAT dummy for that time period either.

3.2 Covariates

We include several sets of covariates on the right-hand side of the equation (1). All variables are treated as time dependent unless said otherwise. The first set consists of macroeconomic variables. Variable YPC is the logarithm of a country's GDP per capita at Purchasing Power Parity (PPP). In view of the early adoption of the VAT in Western Europe, we would expect that wealthy countries are more likely to adopt a VAT, reflecting their more sophisticated tax administrations. Openness ($OPEN$), which is measured as the sum of the GDP shares of goods imports and exports, is likely to have a positive effect on VAT adoption. Intuitively, on average 55 percent of gross VAT revenue is collected at the border (cf. Ebrill et al. 2001), giving open countries an easy tax handle. The share of agriculture in GDP (AGR) captures the notion that VAT introduction is less likely in countries that have a large informal sector. Further, we capture country size by population density ($POPD$), which is expected to have a positive effect on VAT adoption. Finally, revenue (REV), which is measured as the total revenue-to-GDP ratio, is likely to have a negative effect on VAT adoption. To raise public revenue, countries with low revenue ratios would be inclined to adopt a VAT: as Keen and Lockwood (2006) show, the VAT is a 'money machine' because it is a more effective instrument to raise revenue than other consumption taxes. All data on the macroeconomic variables are taken from the World Bank's *World Development Indicators*.

The second set consists of institutional variables. We include a dummy for federal countries (FED) to capture the specific challenges posed by federal systems. The data are taken from Treisman (2008). Keen and Lockwood (2010) argue that taking up a VAT may be less likely in federal systems as they reserve extensive powers over sales taxation to lower levels, making it hard to coordinate tax collection across jurisdictions. We expect this variable to have a negative effect on VAT adoption. Ebrill et al. (2001) argue that the IMF has been an active participant in the spread of the VAT. This participation has taken two main forms: (i) the

[6] According to Shoup (1973), the first consumption-type VAT was introduced in Brazil in January 1967. France introduced a manufacturing-type VAT in 1948, which was extended to cover the retail stage in 1954. This thesis only cover consumption-type VATs.

provision of technical assistance to countries; and (ii) the exercise of influence via lending program conditionality. We capture the influence of the IMF through a dummy variable (*IMF*), which measures whether the country has received financial support from the IMF via the stand-by arrangement or extended fund facility. We expect the IMF to have a positive effect on VAT adoption. Finally, we consider the dummy variable *WAR*, which indicates whether a country has experienced an armed conflict or not in a given year. As tax reform incentives of countries in a war are likely to be less than non-war countries, we expect a negative sign for this variable.

To capture regional effects, we include five regional dummies, that is, Western Hemisphere (*WH*), Middle East and Central Asia (*MECA*), Europe (*EU*), Asia and Pacific (*AP*), and Africa (*AF*). Although there is no constant in our model, including all time dummies and regional dummies would result in a multicollinearity problem with the baseline hazard. In order to avoid this problem, the regional dummy *AF* for the African region is taken as the base case in our analysis. Note that, by including the regional dummies, the baseline hazard is allowed to differ by a constant among various regions (irrespective of time). A more general setting, where the baseline hazard function depends both on time and region in a fully general way as in (4), was analyzed, but the estimates of the baseline hazard were not significantly different from the simpler specification, which is used in the rest of the paper.

Next, because *FED* and the regional dummies hardly change over time, they should be considered as duration independent covariates. Therefore, the set of duration dependent covariates consists of *YPC*, *OPEN*, *AGR*, *POPD*, *WAR*, *IMF*, and *REV* and our duration independent covariates are *FED*, *MECA*, *EU*, *WH*, and *AP*.

We include the spatially lagged dependent variable—which measures strategic interaction among governments—with a two-years lag similar to the other explanatory variables, whereas Keen and Lockwood (2010) estimate the probability of a country adopting a VAT in response to the contemporaneous proportion of neighbors in the region having implemented such a tax. The copycat effect is founded in 'yardstick competition' (Besley and Case, 1995), that is, voters use information on tax systems of neighboring jurisdictions to judge the performance of the politicians of their home jurisdiction.[7] As Keen and Lockwood (2006) demonstrate,

[7]More generally, strategic interaction is observed in countries' decisions whether or not to adopt tax laws. Alm et al. (1993) find evidence of a copycat effect in state lottery enactment in the United States using data for the 1964–1992 period. In addition, Davies and Naughton (2007) and Egger and Larch (2008) provide evidence of increased coalition formation among proximate countries for the case of environmental treaties and

the money machine nature of the VAT leads to an increase in voter welfare. Consequently, rational politicians will mimic the tax setting of neighboring jurisdictions. Hence, we expect the number of VATs of neighboring countries to have a positive effect on the likelihood of a country implementing a VAT.

3.3 MCMC Setup

The analysis is performed based on model described in Section 2. The algorithm was coded in Matlab. We ran two separate MCMC chains with 12,000 iterations with different starting values for each chain. The overlapping trace plots of the parameters show convergence after around 8,000 iterations. In addition, the Gelman-Rubin diagnostics (Gelman and Rubin 1992) also confirm convergence. Therefore, we discarded the first 8,000 iterations as a pre-convergence burn-in period. The last 4,000 samples were used for the posterior analysis in this section. The hyper-parameters used for the prior distributions of the coefficients were set as follows: $b_\theta = 0$ and $T_\theta = I$ (identity matrix). In fact, in the Metropolis-Hastings sampling, we update each parameter in θ separately. Moreover, for the proposed variance of each parameter, a tuning parameter is applied for controlling the acceptance rate to be in the suitable range. Here, we adopt a vague $IG(\sigma_a, \sigma_b)$ prior for σ^2, where σ_a and σ_b are uniform draws from the interval $(1, 2)$. The hyper-parameters of the Gamma distribution for ϕ were set as $(\phi^*, 1)$, where ϕ^* is such that $\varphi(d_{max}/2; \phi^*) = 0.05$, which implies the distances (d) larger than half the maximum observed distance (d_{max}) have low correlation and that the prior for ϕ is centered around the prior guess ϕ^*; see Bastos and Gamerman (2006).

3.4 Results

The summary of estimation results is provided in Table 3. Before discussing and interpreting them, let us first mention some important characteristics of the MCMC estimation procedure related to the data.

Figure A.1 shows the sampling traceplot for all parameters and hyper-parameters in the sample with 92 countries and the contiguity weight matrix. The sampling traceplots of other model specifications are not displayed as they are similar. Short convergence patterns for the parameters are observed along with high autocorrelations in their chains, potentially reflecting

preferential trade agreements, respectively.

Table 3: Results for Sample of 92 Countries with Contiguity Weight Matrix

	No Frailty		Non Spatial		Spatial	
	No Copycat	Copycat	No Copycat	Copycat	No Copycat	Copycat
YPC	-0.1589	-0.1379	-0.1745	-0.1590	-0.1196	-0.1840
	(0.0116)**	(0.0183)**	(0.0178)**	(0.0191)**	(0.0167)**	(0.0138)**
OPEN	-1.5923	-1.6815	-1.5883	-1.8807	-1.7309	-1.7356
	(0.0367)**	(0.0327)**	(0.0685)**	(0.0861)**	(0.0470)**	(0.0549)**
AGR	-1.6523	-1.6275	-1.3966	-1.5817	-0.9609	-1.2069
	(0.0434)**	(0.0516)**	(0.0501)**	(0.0703)**	(0.0918)**	(0.1371)**
POPD	-0.0038	0.0143	-0.0052	0.0051	0.0748	0.1033
	(0.0047)	(0.0044)**	(0.0074)	(0.0071)	(0.0131)**	(0.0095)**
WAR	-0.1978	-0.2305	-0.2215	-0.2996	-0.1118	-0.1189
	(0.0158)**	(0.0153)**	(0.0190)**	(0.0210)**	(0.0278)**	(0.0218)**
IMF	0.7807	0.7662	0.7497	0.7834	0.8165	0.7756
	(0.0163)**	(0.0155)**	(0.0170)**	(0.0153)**	(0.0170)**	(0.0152)**
REV	-0.5196	-0.4406	-0.2494	-0.3310	-0.1479	-0.1105
	(0.0536)**	(0.0730)**	(0.0558)**	(0.0751)**	(0.0718)*	(0.0797)
ρ		0.3478		0.5984		0.3603
		(0.0207)**		(0.0290)**		(0.0262)**
FED	-0.6444	-0.7800	-0.5686	-0.7169	-0.5559	-0.5334
	(0.0191)**	(0.0201)**	(0.0427)**	(0.0498)**	(0.0717)**	(0.0959)**
MECA	-0.6608	-0.6492	-0.7027	-0.6052	-0.3131	-0.2432
	(0.0207)**	(0.0249)**	(0.0455)**	(0.0503)**	(0.1053)**	(0.1019)*
EU	1.4187	1.3928	1.6162	1.6442	1.0563	1.0773
	(0.0338)**	(0.0359)**	(0.0547)**	(0.0538)**	(0.1648)**	(0.0786)**
WH	0.1726	0.1541	0.2791	0.3150	-0.1293	0.1170
	(0.0241)**	(0.0288)**	(0.0409)**	(0.0464)**	(0.1464)	(0.1727)
AP	0.0214	0.0324	-0.0076	0.0409	-0.4143	0.1820
	(0.0180)	(0.0179)	(0.0613)	(0.0510)	(0.1700)*	(0.2153)
$\ln(\lambda_1)$	-1.7469	-1.9048	-1.8846	-1.8574	-1.9106	-1.6593
	(0.0658)**	(0.0988)**	(0.1019)**	(0.0791)**	(0.0722)**	(0.0889)**
$\ln(\lambda_2)$	0.2764	0.0494	0.3406	0.3183	0.5252	0.5636
	(0.0653)**	(0.0967)	(0.0961)**	(0.0680)**	(0.0730)**	(0.0884)**
$\ln(\lambda_3)$	0.6638	0.3760	0.9085	0.8317	1.3091	1.1099
	(0.0619)**	(0.0928)**	(0.0778)**	(0.0620)**	(0.0698)**	(0.0864)**
$\ln(\lambda_4)$	-0.4631	-0.5964	-0.4131	-0.3974	-0.3069	-0.3481
	(0.0344)**	(0.0371)**	(0.0355)**	(0.0335)**	(0.0373)**	(0.0342)**
σ^2			0.3504	0.4736	1.2319	0.7992
ϕ					3252.0	3269.2

Notes: **Significant at 99 percent level; *Significant at 95 percent level.

a low effective sample size (ESS) as reported in Tables A.5 and A.6. Although there is no evidence that the chains of hyper-parameters are converging, this does not cause a concern as the inference on hyper-parameters is not the focus of our analysis.

Tables A.5 and A.6 provide 2.5 percent, 50 percent, 97.5 percent posterior percentiles and posterior means of the parameters of our spatial duration model as discussed in Section 2.1. We also report the ESS and estimated standard errors of the parameters $\hat{V}_{ESS}$. As the the average sample variance would very likely be underestimated due to the positive autocorrelation in the MCMC samples, Kass et al. (1998) use the ESS over the sample variance to be the variance estimator, which is denoted by $\hat{V}_{ESS} = s^2/ESS$, where s^2 is the sample variance and ESS is defined as $ESS = N/\kappa$ using the sample size N, $\kappa = 1+2\sum_{k=1}^{\infty} \rho_k$, and the autocorrelation ρ_k at lag k for each parameter of interest. Empirically, the autocorrelation lag κ can be estimated by using sample autocorrelations estimated from the MCMC chain, cutting off the summation when it drops below 0.1 in magnitude (cf. Roberts, 1996). The last column of these two tables reports the change of relative hazard ratio of VAT adoption of a unit change in a continuous variable or of a change from zero to unity in a dummy.

The estimation results for the complete model are summarized in the last two columns of Table 3, which contains the results using the contiguity spatial weight matrix and 92 countries; the estimated results for both samples with both contiguity and distance weight matrix can be found in Tables A.5 and A.6, respectively. For the sample with 92 countries, all estimated parameters are significant, except for the revenue-to-GDP ratio and two regional dummies (WH and AP). The negative sign of the YPC coefficient shows that less wealthy countries primarily adopt a VAT, which is not in line with expectations. This counterintuitive sign can be explained by the pattern of VAT spread. Recent VAT adopters are less prosperous economies than the early adopters, which were primarily industrialized countries. Indeed, most of the VAT spread over the last twenty years occurred in Africa (see Table 2). In addition, once countries become rich, the composition of the tax mix shifts toward income taxation, making them less dependent on consumption taxation. Another striking result is that openness enters with a significantly negative effect, which coincides with the findings of Desai and Hines (2005) and Keen and Lockwood (2010). For the other variables in the estimation, all results are in line with our expectations. VAT adoption is more likely in countries with a small share of agriculture as the agricultural sector is hard to tax. Countries with a larger population density are also more likely to implement a VAT. Further, if a country experienced an armed conflict, then the relative hazard rates will decrease by around 11 percent (see Table A.5) while keeping the other variables unchanged.

For the institutional effects, there seems to be a significant challenge for federal countries to implement a VAT. The relative hazard rates of VAT adoption for federal systems are about 41 percent less than for non-federal countries (see Tables A.5 and A.6). However, countries with an IMF program have a higher probability than non-program countries, which confirms the significant role of the IMF in the spread of the VAT.

Our findings provide evidence that the VAT has tended to spread in regional bursts. As compared to countries in Africa, countries in Europe are more inclined to implement a VAT while countries in the Middle East and Central Asia have a smaller probability of adopting a VAT. Besides the regional effects, the copycat effect is another locational impact that drives the spread of VAT adoption. This effect is positive and quite robust in our estimation when using the contiguity weight matrix. However, when the distance weight matrix is employed for estimation, the copycat effect in the larger sample is no longer significant as indicated in Table A.6. It is evident that direct international spillovers of the VAT adoption seem to exist only among neighboring countries, which is also found by Keen and Lockwood (2010), who do not use a formal spatial analysis.

Finally, results in Table 3 are also used to check the robustness of the copycat effect for different model specifications. In addition to models with spatial frailties, there are models without frailties and models with non-spatial frailties. Each kind of model is estimated with and without the copycat effect. As reported in the table, the copycat-effect presence is quite robust to model specification.

For the model with spatial frailties, including the spatially weighted lagged dependent variables generally increases the absolute size of the effect of *YPC* and *POPD*. Further, comparing models with and without spatial frailties yields the following observations. The first one is that the coefficient of *POPD* changes from insignificantly negative to significantly positive. The second one is that the effect of the revenue-to-GDP ratio is no longer significant. Finally, although the significance of the baseline hazard rates remains unchanged, their magnitude increases for the second and the third time intervals. Therefore, incorporating country-level unobserved heterogeneity and allowing it to be spatially correlated does have a significant effect on the other estimates. However, inference based on the estimated frailties will not be performed here as it is not the main focus of this paper and left for future research.

4 Conclusions

We have described several approaches to model spatial correlation structures in spatial survival analysis and illustrated the methods with a unique dataset on value added tax (VAT) adoption. More specifically, we examined potential copycat behavior by governments in adopting a VAT. By including the spatially correlated dummy dependent variables in the hazard function of a duration model, we explicitly modeled and estimated this copycat effect rather than imposing that the spatial effects only exist in the frailties. We also allow for region-specific baseline hazards. The adoption of Bayesian methods, implemented via Markov chain Monte Carlo algorithms, enables full posterior inference not only on the effects of the main covariates, but also on the country-level frailties.

We find strong evidence of a copycat effect irrespective whether a contiguity or distance weight matrix is used. The copycat effect is quite robust to model specifications with and without frailties. As compared to countries in Africa, countries in Europe are more inclined to implement a VAT, while countries in the Middle East and Central Asia have a smaller probability of adopting a VAT. The VAT spread occurs in regional bursts. Countries contemplating adopting a VAT may therefore usefully benefit from regionally coordinated technical assistance provided by donor countries.

We have not performed an analysis and inference based on the estimated frailties, which we leave for future work. Finally, further research could usefully consider a continuous baseline hazard function or nonparametric alternatives to our piecewise baseline hazard function.

Acknowledgements

The authors would like to thank the IMF's Tax Policy Division for making available the data on revenues and VAT rates, the conference participants at the VI World Conference of the Spatial Econometrics Association (Salvador, Brazil, July 2012), and seminar participants at Tilburg University.

References

Alm J, McKee M, Skidmore M. 1993. Fiscal pressure, tax competition, and the introduction of state lotteries. *National Tax Journal* **46**: 463–76.

Banerjee S, Wall MM, Carlin BP. 2003. Frailty modeling for spatially correlated survival data, with application to infant mortality in minnesota. *Biostatistics* **4**: 123–142.

Bastos LS, Gamerman D. 2006. Dynamic survival models with spatial frailty. *Lifetime Data Analysis* **12**: 441–460.

Besag J. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)* **36**: 192–236.

Besley T, Case A. 1995. Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition. *The American Economic Review* **85**: 25–45.

Brockmeyer A. 2010. Does IMF conditionality work? Evidence from VAT adoption. Mimeo, London School of Economics, London.

Carlin BP, Hodges JS. 1999. Hierarchical proportional hazards regression models for highly stratified data. *Biometrics* **55**: 1162–1170.

Clayton DG. 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**: 141–151.

Cox DR. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)* **34**: 187–220.

Davies R, Naughton HT. 2007. Cooperation in environmental policy: A spatial approach. mimeo, University of Oregon, Oregon.

Desai MA, Hines JR. 2005. Value-added taxes and international trade: The evidence. mimeo, University of Michigan, Michigan.

Dreher A. 2006. IMF and economic growth: The effects of programs, loans, and compliance with conditionality. *World Development* **34**: 769–788.

Ebrill L, Keen M, Bodin JP, Summers V. 2001. *The Modern VAT*. Washington, D.C.: International Monetary Fund, Washington D.C.

Egger P, Larch M. 2008. Independent preferential trade agreement membership: An empirical analysis. *Journal of International Economics* **76**: 384–399.

Gamerman D. 1991. Dynamic Bayesian models for survival data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **40**: 63–79.

Gelman A, Carlin JB, Stern HS, Rubin DB. 2004. *Bayesian Data Analysis*. London: Chapman and Hall, 2nd edition.

Gelman A, Rubin DB. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* **7**: 457–472.

Geman SGEM, Geman DGEM. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell* **6**: 721–741.

Gupta S. 1991. Stochastic models of interpurchase time with time-dependent covariates. *Journal of Marketing Research* **28**: 1–15.

Hastings WK. 1970. Monte Carlo sampling methods using markov chains and their applications. *Biometrika* **57**: 97–109.

Hemming K, Shaw J. 2002. A parametric dynamic survival model applied to breast cancer survival times. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **51**: 421–435.

Kass RE, Carlin BP, Gelman A, Neal RM. 1998. Markov chain Monte Carlo in practice: A roundtable discussion. *The American Statistician* **52**: 93–100.

Keen M, Lockwood B. 2006. Is the vat a money machine? *National Tax Journal* **LIX**: 905–928.

Keen M, Lockwood B. 2010. The value added tax: Its causes and consequences. *Journal of Development Economics* **92**: 138–151.

Li Y, Ryan L. 2002. Modeling spatial survival data using semiparametric frailty models. *Biometrics* **58**: 287–297.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**: 1087–1092.

Nellor D. 1987. The effect of the value-added tax on the tax ratio. IMF Working Paper no. 87/47, International Monetary Fund, Washington, D.C.

Ripley BD. 1981. *Spatial Statitics*. New York: John Wiley.

Roberts G. 1996. Markov chain concepts related to sampling algorithms. In W R Gilks DJS, Richardson S (eds.) *MCMC in Practice*. London: Chapman and Hall, 45–57.

Roberts GO, Gelman A, Gilks WR. 1997. Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability* **7**: 110–120.

Roberts GO, Rosenthal JS. 2001. Optimal scaling for various metropolis-hastings algorithms. *Statistical Science* **16**: 351–367.

Shoup CS. 1973. *The Value-Added Tax*. Athens: Center of Planning and Economic Research.

Treisman D. 2008. Decentralization Dataset 2008. Available at: http://www.sscnet.ucla.edu/polisci/faculty/treisman/. Date accessed: April 2010, University of California.

Vaupel JW, Manton KG, Stallard E. 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**: 439–454.

APPENDIX

Table A.1: Countries in the Sample

Afghanistan	Grenada	Pakistan
Australia	Guatemala	Papua New Guinea
Austria	Guinea	Paraguay
Bahamas	Honduras	Peru
Bahrain	Iceland	Philippines
Bangladesh	India	Rwanda
Barbados	Indonesia	Senegal
Belize	Iran	Seychelles
Benin	Ireland	South Africa
Bhutan	Italy	Spain
Botswana	Japan	Sri Lanka
Burkina Faso	Jordan	St. Kitts and Nevis
Cambodia	Kenya	St. Vincent and the Grenadines
Cameroon	Korea	Sudan
Canada	Kuwait	Swaziland
Cape Verde	Lao People's Democratic Republic	Switzerland
Central African Republic	Lebanon	Syrian Arab Republic
Chile	Lesotho	Turkey
China	Morocco	Tunisia
Congo (Democratic Republic)	Madagascar	Uganda
Congo (Republic of)	Malaysia	United Arab Emirates
Dominican Republic	Maldives	United Kingdom
Egypt	Mali	United States
El Salvador	Mauritius	Vanuatu
Estonia	Mexico	Venezuela
Ethiopia	Mongolia	Vietnam
Fiji	Namibia	Zambia
Finland	Nepal	Zimbabwe
Gambia	New Zealand	
Ghana	Niger	
Greece	Oman	

Notes: Based on a sample of 92 countries. Countries from the Former Soviet Union (i.e., Armenia, Azerbaijan, Belarus, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Moldova, Russia, Tajikistan, and Ukraine) and Eastern Europe and the Balkans (i.e., Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Hungary, Poland, Romania, Slovak Republic, Slovenia) are excluded from the sample.

Table A.2: Descriptive Statistics

Variable	Observations	Mean	St.Dev.	Min	Max
DUR	92	25.326	10.182	2.000	38.000
YPC	3,680	8.202	1.319	4.767	10.939
$OPEN$	3,362	0.712	0.406	0.053	3.754
AGR	3,162	0.197	0.406	0.002	0.743
$POPD$	3,543	1.082	1.536	0.008	9.884
WAR	3,680	0.198	0.400	0	1.000
IMF	3,680	0.083	0.277	0	1.000
REV	2,089	0.222	0.119	0.002	0.823
FED	3,680	0.160	0.367	0	1.000
$MECA$	3,680	0.152	0.360	0	1.000
EU	3,680	0.109	0.311	0	1.000
WH	3,680	0.196	0.397	0	1.000
AP	3,680	0.240	0.427	0	1.000
AF	3,680	0.304	0.460	0	1.000
y_t	3,680	0.817	0.388	0	1.000

Notes: The variables are described in Table A.3.

Table A.3: Data Description and Sources

Variable	Definition	Sources
$DUR(t)$	Time spell before VAT adoption (starting from 1970)	Ebrill et al. (2001), internal database of the IMF's Fiscal Affairs Department (Tax Policy Division), and IBFD's *Tax News Service*
YPC	The logarithm of GDP per capita at PPP (in thousands of 2005 US dollars)	World Bank (2011), *World Development Indicators* (WDI)
$OPEN$	The sum of goods imports and exports as share of GDP	World Bank (2011), WDI
AGR	Share of agriculture in GDP	World Bank (2011), WDI
$POPD$	Population density (in hundreds of people per square kilometers of land area)	World Bank (2011), WDI
WAR	Dummy variable that takes on a value of unity if country is in armed conflict at time t and zero otherwise	Comprehensive Study of Civil War (CSCW), *Annual Report 2009*, available at: http://www.prio.no/CSCW
IMF	Dummy variable that takes the value of one if the country has concluded a Stand-By Arrangement (SBA), Extended Fund Facility (EFF), or Poverty Reduction and Growth Facility (PRGF) with the IMF	Dreher (2006)
REV	General government revenue (including grants) as share of GDP, 1975–2007.. REV is based on series R^A and R^B. The procedure to derive REV from R^A and R^B is described in Section 3.2	IMF, The series R^A is derived from *Government Finance Statistics*, *IMF Staff Reports*, *IMF Selected Issues Papers*. The series R^B is obtained from an Internal database of the IMF's Fiscal Affairs Department (Tax Policy Division)
FED	Dummy variable that takes on a value of unity if country has a federal structure and zero otherwise	Treisman (2008)
$MECA$, EU, WH, AP, AF	Regional dummies for Middle East and Central Asia, Europe, Western Hemisphere, Asia and Pacific, and Africa, respectively	
δ	Dummy variable that takes on a value of unity in the absence of censoring and zero for right censoring	Authors' calculations
y_t	Dummy variable that takes on a value of unity if the country has a VAT in year t and zero otherwise	Authors' calculations. See duration source
w_{ij}^C	$w_{ij}^C = b_{ij}/\sum_{i=1}^N b_{ij}$ for $i \neq j$ and $w_{ij}^C = 0$ for $i = j$, where b_{ij} is a border dummy which equals one if countries i and j share a common border and zero otherwise.	Authors' calculations
w_{ij}^D	$w_{ij}^D = \begin{cases} \frac{1}{d_{ij}^2}/\sum_{j=1}^N \frac{1}{d_{ij}^2} > 0 & \text{for } i \neq j \\ 0 & \text{for } i = j \end{cases}$, where d_{ij} reflects the geographical distance between the largest cities of countries i and j, which is computed as the great circle distance given latitude and longitude.	Authors' calculations based on distance data from CEPII, see http://www.cepii.fr/

Notes: Missing data in YPC, $OPEN$, AGR, $POPD$, and REV are interpolated. See Section 3.2.

Table A.4: Missing Data in the Subsample

Variable	Observations	Missing	Freq. Missing
YPC	4,320	80	1.85
OPEN	4,320	495	11.46
AGR	4,320	801	18.54
POPD	4,320	265	6.13
WAR	4,320	0	0
IMF	4,360	0	0
REV	4,320	1,224	28.33
R^A	4,320	2,132	49.35
R^B	4,320	2,340	54.17
FED	4,320	0	0
MECA	4,320	0	0
EU	4,320	0	0
WH	4,320	0	0
AP	4,320	0	0

Notes: The variables are described in Table A.3.

Figure A.1: MCMC Sampling for the Parameters

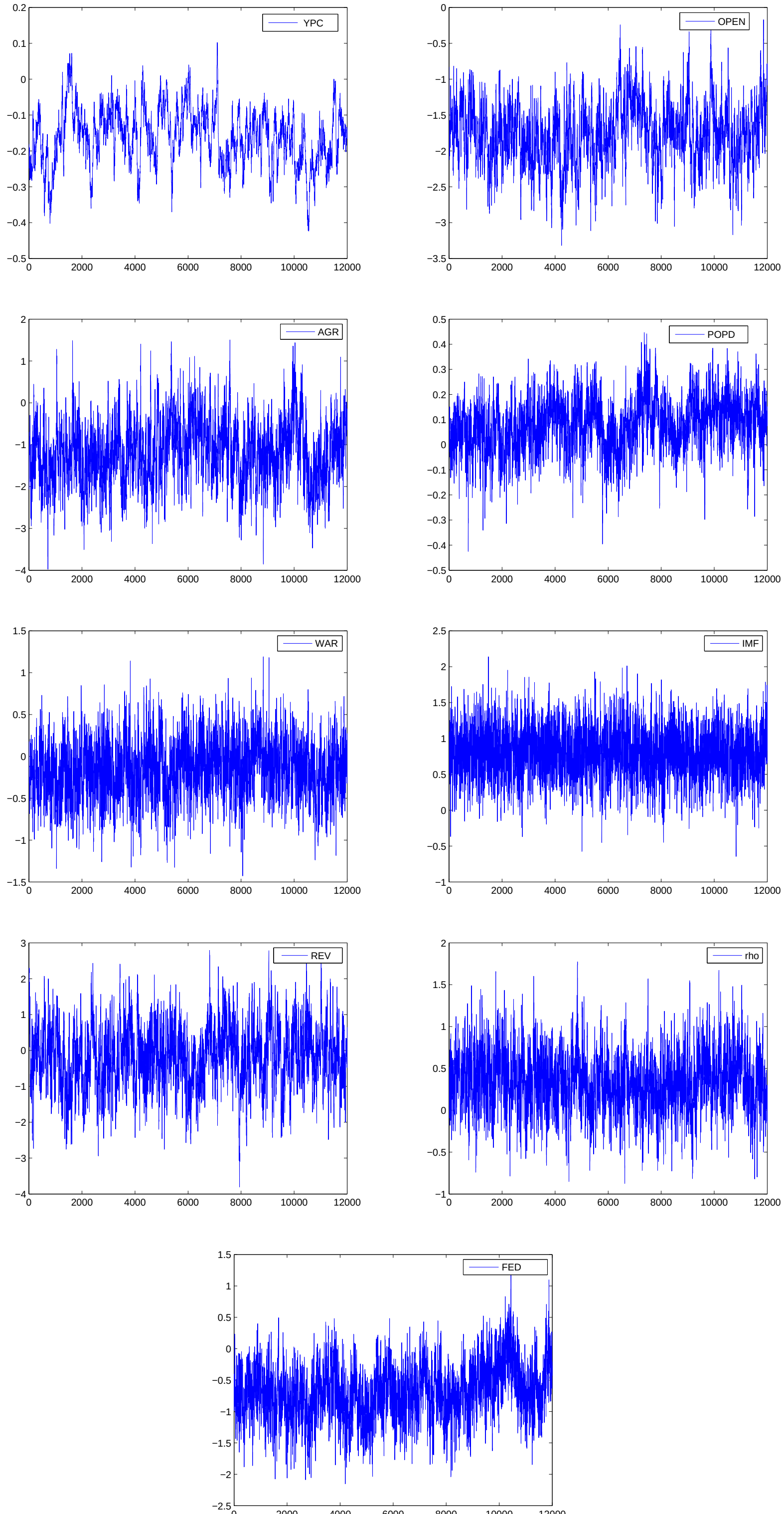


Figure A.1: MCMC Sampling for the Parameters (Continued)

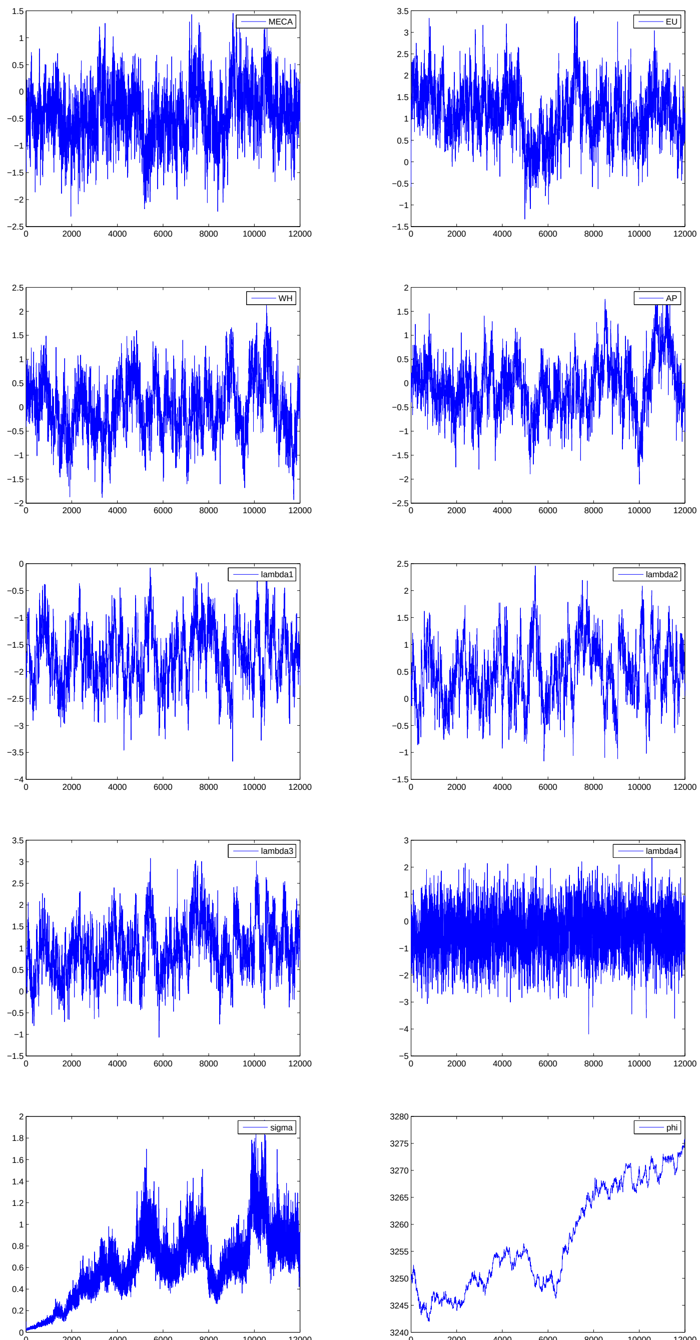


Table A.5: Results with Contiguity Weight Matrix

Covariates	Mean	2.5%	50%	97.5%	ESS	$\sqrt{\hat{V}_{ESS}}$	exp(β)-1
			Sample with 52 countries				
YPC	-0.1261	-0.3136	-0.1222	0.0778	23.4263	(0.0210)**	-0.1185
OPEN	-1.1475	-2.0123	-1.1891	-0.1037	52.9381	(0.0660)**	-0.6826
AGR	-0.7721	-2.5176	-0.7952	1.0342	31.7765	(0.1608)**	-0.5380
POPD	0.0601	-0.2719	0.0669	0.3750	81.3898	(0.0181)**	0.0619
WAR	-0.4915	-1.4276	-0.4657	0.2965	323.0010	(0.0237)**	-0.3883
IMF	0.7786	-0.0052	0.7797	1.5351	301.9281	(0.0224)**	1.1783
REV	0.0593	-1.8549	0.0696	1.8102	96.2077	(0.0956)	0.0611
ρ	0.3073	-0.5338	0.2924	1.1747	81.4413	(0.0484)**	0.3597
FED	-0.3431	-1.2184	-0.3520	0.6413	87.9173	(0.0519)**	-0.2904
MECA	-0.3529	-1.8573	-0.3441	1.0378	47.8989	(0.1079)**	-0.2973
EU	1.0827	-0.2901	1.0462	2.4984	41.1633	(0.1134)**	1.9527
WH	-0.1605	-1.5313	-0.1760	1.2318	29.1298	(0.1309)	-0.1483
AP	-0.4170	-1.6984	-0.4309	0.8843	21.8029	(0.1496)**	-0.3410
$\ln(\lambda_1)$	-2.0986	-3.1779	-2.1141	-0.9706	49.1720	(0.0837)**	
$\ln(\lambda_2)$	0.4555	-0.6477	0.4489	1.6423	41.7052	(0.0907)**	
$\ln(\lambda_3)$	0.9892	-0.2393	0.9817	2.3842	63.9082	(0.0842)**	
$\ln(\lambda_4)$	-0.1903	-2.0563	-0.1811	1.5987	504.3555	(0.0416)**	
σ^2	1.8281	0.7846	1.6627	3.6896			
ϕ	3251.5	3246.0	3251.5	3256.1			
			Sample with 92 countries				
YPC	-0.1840	-0.3395	-0.1798	-0.0580	26.6026	(0.0138)**	-0.1681
OPEN	-1.7356	-2.6137	-1.7263	-0.8759	59.1524	(0.0549)**	-0.8237
AGR	-1.2069	-2.6144	-1.2138	0.3078	30.0958	(0.1371)**	-0.7009
POPD	0.1033	-0.0891	0.1081	0.2861	95.5184	(0.0095)**	0.1088
WAR	-0.1189	-0.7537	-0.1169	0.4668	213.7671	(0.0218)**	-0.1121
IMF	0.7756	0.1040	0.7659	1.4011	465.3071	(0.0152)**	1.1720
REV	-0.1105	-1.9247	-0.0840	1.5591	116.1541	(0.0797)	-0.1047
ρ	0.3603	-0.3329	0.3634	1.0723	182.7501	(0.0262)**	0.4338
FED	-0.5334	-1.4179	-0.5419	0.3795	22.2999	(0.0959)**	-0.4134
MECA	-0.2432	-1.3578	-0.2451	0.7971	28.0986	(0.1019)*	-0.2159
EU	1.0773	-0.0480	1.0464	2.1991	51.9621	(0.0786)**	1.9367
WH	0.1170	-1.0733	0.1043	1.4404	15.6244	(0.1727)	0.1241
AP	0.1820	-1.3263	0.1862	1.4928	10.2884	(0.2153)	0.1996
$\ln(\lambda_1)$	-1.6593	-2.7843	-1.6526	-0.6604	38.7612	(0.0889)**	
$\ln(\lambda_2)$	0.5636	-0.5319	0.5767	1.6023	37.9608	(0.0884)**	
$\ln(\lambda_3)$	1.1099	0.0364	1.1020	2.3653	46.8258	(0.0864)**	
$\ln(\lambda_4)$	-0.3481	-2.0139	-0.3037	1.2839	603.1201	(0.0342)**	
σ^2	0.7992	0.3607	0.7628	1.4357			
ϕ	3269.2	3264.9	3268.7	3274.2			

Notes: **Significant at 99 percent; *Significant at 95 percent.

Table A.6: Results with Distance Weight Matrix

Covariates	Mean	2.5%	50%	97.5%	ESS	$\sqrt{\hat{V}_{ESS}}$	$exp(\beta)$-1
			Sample with 52 countries				
YPC	-0.1206	-0.3023	-0.1183	0.0757	38.6403	(0.0154)**	-0.1136
OPEN	-1.0906	-2.1804	-1.0574	-0.0959	43.5671	(0.0829)**	-0.6640
AGR	-0.7038	-2.2988	-0.7174	0.9308	119.0034	(0.0764)**	-0.5053
POPD	-0.0600	-0.5195	-0.0539	0.3220	31.8803	(0.0364)	-0.0582
WAR	-0.5163	-1.4617	-0.5064	0.3081	86.7805	(0.0480)**	-0.4033
IMF	0.8113	-0.0472	0.8170	1.5696	301.1268	(0.0235)**	1.2508
REV	0.0475	-2.0278	0.0715	1.8073	203.7966	(0.0651)	0.0486
ρ	0.5372	-0.6530	0.5535	1.7016	116.5859	(0.0559)**	0.7111
FED	-0.3185	-1.3454	-0.3090	0.7075	45.7260	(0.0773)**	-0.2728
MECA	-0.6494	-2.0412	-0.6298	0.5492	104.4693	(0.0631)**	-0.4777
EU	0.9729	-0.4717	1.0053	2.2948	32.5642	(0.1203)**	1.6457
WH	0.3611	-0.7952	0.3097	1.7041	27.6513	(0.1249)**	0.4350
AP	-0.5340	-2.0175	-0.5153	0.7503	26.7544	(0.1404)**	-0.4138
$\ln(\lambda_1)$	-2.0621	-3.3407	-2.0802	-0.7457	41.7612	(0.1042)**	
$\ln(\lambda_2)$	0.6361	-0.4436	0.6325	1.7397	35.5095	(0.0937)**	
$\ln(\lambda_3)$	0.9755	-0.3530	0.9785	2.2926	52.5243	(0.0903)**	
$\ln(\lambda_4)$	-0.2258	-2.0292	-0.2246	1.6007	606.1145	(0.0374)**	
σ^2	2.3996	0.5082	2.2280	5.6445			
ϕ	3250.8	3248.0	3250.7	3253.2			
			Sample with 92 countries				
YPC	-0.1406	-0.3187	-0.1422	0.0404	7.6199	(0.0362)**	-0.1312
OPEN	-1.7065	-2.5780	-1.7351	-0.7733	18.9087	(0.1038)**	-0.8185
AGR	-0.8980	-2.2897	-0.8871	0.5362	82.1179	(0.0828)**	-0.5926
POPD	0.1013	-0.0940	0.1064	0.2729	97.1077	(0.0097)**	0.1067
WAR	-0.1340	-0.8107	-0.1295	0.5159	106.5908	(0.0327)**	-0.1254
IMF	0.7706	0.0307	0.7859	1.4169	421.9398	(0.0170)**	1.1611
REV	-0.0327	-1.8664	-0.0419	1.8057	149.8653	(0.0706)	-0.0322
ρ	0.0143	-0.9938	0.0260	0.9145	114.2905	(0.0458)	0.0144
FED	-0.5283	-1.3286	-0.5184	0.2676	109.7729	(0.0387)**	-0.4104
MECA	-0.3799	-1.5374	-0.3885	0.8938	22.9936	(0.1317)**	-0.3161
EU	1.1043	-0.1351	1.0802	2.4957	25.7005	(0.1337)**	2.0172
WH	0.0618	-1.3060	0.0716	1.3313	23.1900	(0.1355)	0.0637
AP	0.1177	-1.0483	0.1540	1.1715	19.3569	(0.1397)	0.1249
$\ln(\lambda_1)$	-1.9062	-3.0687	-1.9276	-0.7408	30.7631	(0.1061)**	
$\ln(\lambda_2)$	0.5220	-0.5098	0.5245	1.6349	38.4572	(0.0868)**	
$\ln(\lambda_3)$	1.2690	0.1345	1.2663	2.4676	58.4051	(0.0745)**	
$\ln(\lambda_4)$	-0.2974	-2.0050	-0.2958	1.3639	638.2705	(0.0342)**	
σ^2	1.2828	0.6392	1.2111	2.2577			
ϕ	3251.7	3249.69	3251.9	3253.3			

Notes: **Significant at 99 percent; *Significant at 95 percent.